PATENT ABSTRACTS OF JAPAN

(11)Publication number:

11-232304

(43)Date of publication of application: 27.08.1999

(51)Int.Cl.

G06F 17/30 G06F 13/00 G06F 13/00 G06F 17/27 H04L 12/54 H04L 12/58

(21)Application number : 10-042896

(71)Applicant: CASIO COMPUT CO LTD

(22)Date of filing:

09.02.1998

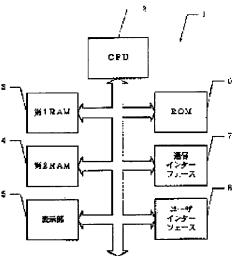
(72)Inventor: KAFUKU SHIGERU

(54) DEVICE FOR JUDGING CONTENTS OF SENTENCE AND ELECTRONIC MAIL DEVICE USING THE JUDGING DEVICE

(57)Abstract:

PROBLEM TO BE SOLVED: To make it possible to judge a sentence including a specific word and to previously prevent the reading or distribution of the sentence by providing the judging device with a 1st judging means for judging whether an extracted word is registered in a word dictionary or not and a 2nd judging means for judging whether the contents of a sentence to be analyzed include specific idea or not.

SOLUTION: A CPU 2 executes a required program under an operating system and attains respective functions of a reading means, an extraction means, the 1st judging means, a counter, and the 2nd judging means by executing an electronic (E) mail program to be one of plural programs. A 1st RAM 3 provides a working storage space for executing a program. A 2nd RAM 4 provides a storage space for storing an incoming or transmitting E mail especially at the time of executing an E mail program. A ROM 6 stores 'word dictionary' necessary for the execution of the E mail program or the like. A display part 5 displays the execution result of a program or the like.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

Searching PAJ 2/2 ページ

[Number of appeal against examiner's decision of rejection]
[Date of requesting appeal against examiner's

decision of rejection]

[Date of extinction of right]

Reference 1: Japanese Patent Laid Open No. 11-232304 Paragraph [0010]

On the other hand, when a file is attached to e-mail, a text data extracted from the attached file (S4) is divided into a word by using such as a morphological analysis **, for example, Then, each word is subjected to be matched with a word dictionary 10 in order to count words coincided with the dictionary. Comparing the counted value with reference value, the file is allowed for displaying (S7) when the value does not exceed the reference value (a result for determination; No or False), and the file is rejected for displaying or alarmed when the value exceeds the reference value (a result for determination; Yes or True).

** morphological analysis; it is a method for analysis developed in the educational field called morphology wherein a minimum word unit with meaning (morpheme) and an inner structure and a format of a word is set for an object for research. It is especially used for dividing a character string into a word in a unit for converting Japanese words of the word processor.

Paragraph [0012]

In this embodiment, a registrar of the word dictionary 10 and a recipient of email is another person, and their sense is not the same depending on the circumstances (when even a registrar feels unpleasant for the word, a recipient does not necessarily feel unpleasant). In such a case, the recipient may edit the contents of the word dictionary by himself.

Paragraph [0016]

Embodiments of the present invention are not limited thereto. Various modifications are contained within the intention of the present invention. For, example, a plurality of reference values each having a different reference value may be arranged so that levels of rejection and alarming may be adjusted. Further, although unpleasant e-mail is rejected for receiving and sending as described above, it may be set to be illegible or matched words are changed to turned letters or replaced with another language codes. ... omitted....

(19)日本国特許庁 (JP)

(12) 公開特許公報(A)

(11)特許出願公開番号

特開平11-232304

(43)公開日 平成11年(1999)8月27日

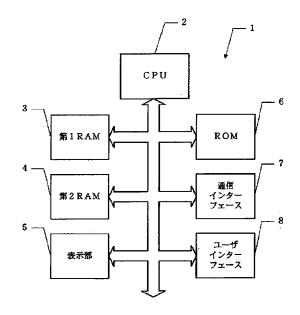
(51) Int.Cl. ⁶		識別記号	FΙ						
G06F	17/30			G 0	6 F	15/40		370A	
	13/00	351				13/00		351G	
		3 5 4						354D	
	17/27					15/20		550A	
H04L	12/54					15/403		340A	
			客查請求	未請求	請求	項の数7	FD	(全 7 頁)	最終頁に続く
(21)出願番号		特顧平10-4289 6		(71) 出願人 000001443					
						カシオ	計算機	株式会社	
(22)出願日		平成10年(1998) 2月9日		東京都渋谷区本町1丁目6番2号					
				(72)	(72)発明者		滋		
						東京都	羽村市	榮町3丁目2 :	番1号 カシオ
						計算機	株式会	社羽村技術セ	ンター内
				(74)	代理人	、弁理士	鹿嶋	英實	

(54) 【発明の名称】 文章内容判定装置及びそれを用いた電子メール装置

(57)【要約】

【課題】 主観上特定の思想に属する単語又は思想を想起し得る単語(以下、特定単語)を含む文章を判定でき、特に電子メールに適用して好適な文章内容判定装置を提供する。

【解決手段】 特定単語を登録した単語辞書、解析対象の文章を読み取る読み取り手段、文章から単語を抽出する抽出手段、抽出された単語が単語辞書に登録されているか否かを判定する第1判定手段、第1判定手段による判定結果が否でない場合の頻度に対応して値を変化させるカウンタ、及びカウンタの値が基準値に達したとき、解析対象の文章の内容が特定の思想を含むものであることを判定する第2判定手段を備える。特定単語を含む文章を判定でき、かかる文章の閲覧や他者への配布を未然に防止できる。



【特許請求の範囲】

【請求項1】 主観上特定の思想に属する単語又は該思 想を想起し得る単語を登録した単語辞書と、

解析対象の文章を読み取る読み取り手段と、

前記文章から単語を抽出する抽出手段と、

前記抽出された単語が前記単語辞書に登録されているか 否かを判定する第1判定手段と、

前記第1判定手段による判定結果が否でない場合の頻度 に対応して値を変化させるカウンタと、

前記カウンタの値が基準値に達したとき、前記解析対象 10 の文章の内容が前記特定の思想を含むものであることを 判定する第2判定手段と、

を備えたことを特徴とする文章内容判定装置。

【請求項2】 主観上特定の思想に属する単語又は該思 想を想起し得る単語を、その出現頻度を表す係数ととも に登録した単語辞書と、

解析対象の文章を読み取る読み取り手段と、

前記文章から単語を抽出する抽出手段と、

前記抽出された単語が前記単語辞書に登録されているか 否かを判定する第1判定手段と、

前記第1判定手段による判定結果が否でない場合の頻度 と該場合における判定対象単語に関連する前記係数とに 対応して値を変化させるカウンタと、

前記カウンタの値が基準値に達したとき、前記解析対象 の文章の内容が前記特定の思想を含むものであることを 判定する第2判定手段と、

を備えたことを特徴とする文章内容判定装置。

【請求項3】 前記第2判定手段は、前記カウンタの値 を前記文章の文字数で正規化した後、該正規化後のカウ 内容が前記特定の思想を含むものであることを判定する ことを特徴とする請求項1又は請求項2記載の文章内容 判定装置。

【請求項4】 請求項1又は請求項2記載の単語辞書、 読み取り手段、抽出手段、第1判定手段、カウンタ及び 第2判定手段を備え、

電子メール又は該電子メールの添付ファイルに含まれる テキストデータを解析対象の文章とするとともに、

第2 判定手段の判定結果が真の場合に、該電子メールの 送信、受信又は判読可能な形式での出力を拒否し、若し 40 くは警告を発することを特徴とする電子メール装置。

【請求項5】 請求項4記載の電子メール装置におい て、前記第2判定手段は、値の異なる複数の基準値を用 いて前記解析対象の文章の内容が前記特定の思想を含む ものであることを多段階に判定し、各段階の判定結果に 応じて前記拒否又は警告のレベルを切り替えることを特 徴とする電子メール装置。

【請求項6】 請求項4又は請求項5記載の電子メール 装置において、前記特定の思想を含むものであると判定 の差出人からの電子メールの受信又は表示を拒否すると とを特徴とする電子メール装置。

【請求項7】 請求項1又は請求項2記載の単語辞書、 読み取り手段、抽出手段、第1判定手段、カウンタ及び 第2判定手段を実現するためのプログラムを格納したと とを特徴とする記録媒体。

【発明の詳細な説明】

[0001]

【発明の属する技術分野】本発明は、文章内容判定装置 及びそれを用いた電子メール装置に関し、詳しくは、主 観上好ましくない思想(注1)を内容に含むテキストデ ータの閲覧若しくは他者への転送を未然に防止できる新 規な文章内容判定装置及びそれを用いた電子メール装置 に関する。

注1:主観上好ましくない思想とは、人の情感に働きか けてその人に一定の行動を起こすきっかけを与える様々 な思想のうち、少なくともある人にとっては不快と感じ る思想のことを言う。

[0002]

【従来の技術】電子メールは、電話やファクシミリのよ 20 うに受取人の端末を特定しなくてもよく、しかも、受取 人は自分の好きな時間に読むことができる点で、既存の 情報伝達手段にない優れた長所を持っており、公私を問 わず欠かせない手段の一つになっている。

[0003]

【発明が解決しようとする課題】ところで、電子メール は、受取人のアドレス (インターネットメールの場合は E-mailアドレス) さえ分かれば、その人のメール ボックス(メールサーバ上に設けられた受取人専用のフ ンタの値が基準値に達したとき、前記解析対象の文章の 30 ォルダ)に誰でも自由に投函することができる。このこ と自体は、電子メールの性格からして何らの不都合もな いが、時によっては、そのメールボックスの所有者や会 社などを誹謗中傷し、又は社会通念上有害な内容の電子 メールが投函されることがある。このような場合、受取 人は、それらの電子メールを一つ一つ削除することにな るが、当然ながら、その間不快感を伴うという問題点が ある。また、かかる電子メールが大量に投函された場合 には、削除に手間がかかるばかりか、本来の業務に支障 を与えかねないという問題点がある。

> 【0004】そこで本発明は、主観上特定の思想に属す る単語又は該思想を想起し得る単語を含む文章を判定で き、特に電子メールに適用して好適な文章内容判定装置 の提供を目的とする。

[0005]

【課題を解決するための手段】請求項1記載の発明に係 る文章内容判定装置は、主観上特定の思想に属する単語 又は該思想を想起し得る単語を登録した単語辞書と、解 析対象の文章を読み取る読み取り手段と、前記文章から 単語を抽出する抽出手段と、前記抽出された単語が前記 された電子メールの差出人の情報を保持しておき、同一 50 単語辞書に登録されているか否かを判定する第1判定手

段と、前記第1判定手段による判定結果が否でない場合 の頻度に対応して値を変化させるカウンタと、前記カウ ンタの値が基準値に達したとき、前記解析対象の文章の 内容が前記特定の思想を含むものであることを判定する 第2判定手段と、を備えたことを特徴とする。請求項2 記載の発明に係る文章内容判定装置は、主観上特定の思 想に属する単語又は該思想を想起し得る単語を、その出 現頻度を表す係数とともに登録した単語辞書と、解析対 象の文章を読み取る読み取り手段と、前記文章から単語 を抽出する抽出手段と、前記抽出された単語が前記単語 10 辞書に登録されているか否かを判定する第1判定手段 と、前記第1判定手段による判定結果が否でない場合の 頻度と該場合における判定対象単語に関連する前記係数 とに対応して値を変化させるカウンタと、前記カウンタ の値が基準値に達したとき、前記解析対象の文章の内容 が前記特定の思想を含むものであることを判定する第2 判定手段と、を備えたことを特徴とする。請求項3記載 の発明に係る文章内容判定装置は、請求項1又は請求項 2記載の発明において、前記第2判定手段は、前記カウ ンタの値を前記文章の文字数で正規化した後、該正規化 20 後のカウンタの値が基準値に達したとき、前記解析対象 の文章の内容が前記特定の思想を含むものであることを 判定することを特徴とする。請求項4記載の発明に係る 電子メール装置は、請求項1又は請求項2記載の単語辞 書、読み取り手段、抽出手段、第1判定手段、カウンタ 及び第2判定手段を備え、電子メール又は該電子メール の添付ファイルに含まれるテキストデータを解析対象の 文章とするとともに、第2判定手段の判定結果が真の場 合に、該電子メールの送信、受信又は判読可能な形式で の出力を拒否し、若しくは警告を発することを特徴とす 30 る。請求項5記載の発明に係る電子メール装置は、請求 項4記載の電子メール装置において、前記第2判定手段 は、値の異なる複数の基準値を用いて前記解析対象の文 章の内容が前記特定の思想を含むものであることを多段 階に判定し、各段階の判定結果に応じて前記拒否又は警 告のレベルを切り替えることを特徴とする。請求項6記 載の発明に係る電子メール装置は、請求項4又は請求項 5記載の電子メール装置において、前記特定の思想を含 むものであると判定された電子メールの差出人の情報を 保持しておき、同一の差出人からの電子メールの受信又 40 は表示を拒否することを特徴とする。請求項7記載の発 明に係る記録媒体は、請求項1又は請求項2記載の単語 辞書、読み取り手段、抽出手段、第1判定手段、カウン タ及び第2判定手段を実現するためのプログラムを格納 したことを特徴とする。

[0006]

【発明の実施の形態】以下、本発明の実施の形態を図面 を参照しながら説明する。図1は、LAN、WAN又は インターネットなどに接続して電子メールの送受信を行 装置1は、CPU2、第1RAM3、第2RAM4、表 示部5、ROM6、通信インターフェース7及びユーザ インターフェース8などを備え、これら各部をバス9で つないで構成している。

【0007】ととで、各部の主要な機能を説明すると、 CPU2はオペレーティングシステムの下で所要のプロ グラムを実行するものであり、そのプログラムの一つで ある後述の電子メールプログラムを実行することによっ て、発明の要旨に記載の読み取り手段、抽出手段、第1 判定手段、カウンタ及び第2判定手段の各機能を実現す るものである。また、第1RAM3はブログラムを実行 するための作業用の記憶空間を与えるもの、第2RAM 4は特に電子メールプログラムを実行する際の着信又は 送信電子メールを保存する記憶空間を与えるものであ り、これらの第1RAM3と第2RAM4は必ずしも物 理的に分離している必要はない。記憶空間として独立し ていればよい。

【0008】ROM6は電子メールプログラムなどの実 行に必要な各種バラメータを恒久的に保持するものであ る。特に、本実施の形態においては、後述する「単語辞 書」を保持するものである。表示部5はプログラムの実 行結果などを表示するものであり、特に本実施の形態に おいては、電子メールの作成、受信及び送信(返信や転 送を含む)に必要な様々な表示を行うものである。通信 インターフェース7はLAN、WAN又はインターネッ トなどに接続して電子メールなどの電子データを送受信 するものであり、ユーザーインターフェース8はキーボ ード、マウス又はタッチパネルなどの入力機器とのイン ターフェースをとるものである。

【0009】図2はCPU2で実行する電子メールプロ グラムのフローチャートであり、特に、電子メールの受 信部分を示す概略的なフローチャートである。この図に おいて、電子メールを受信(S1)すると、まず、その 電子メールの添付ファイルの有無を判定(S2)し、そ して、添付ファイルがない場合は電子メールを所要の形 式に加工した後、表示部5に送って表示(S3)する。 【0010】一方、添付ファイルがある場合は、その添 付ファイルからテキストデータを取り出し(S4)、例 えば、形態素解析(注2)などの手法を用いてテキスト データを単語に分解した後、単語ごとにROM6の単語 辞書10とマッチング(S5)をとり、一致した数をカ ウントする。そして、このカウント値と基準値とを比較 (S6) し、基準値を超えていなければ(判定結果: N O又は偽 "False") 添付ファイルの表示を許容(S7) し、超えている場合(判定結果: YES又は真"Tru e") は添付ファイルの表示を拒否(S8) 又は警告す

注2:形態素解析(morphological analysis)とは、意 味を持つ最小の言語単位(形態素)や語(Word)の内部 う電子メール装置のシステム構成図であり、電子メール 50 構造及び形式を研究対象とする形態論(又は語形論)と

呼ばれる学問分野で発達した解析手法であり、特に、ワ ードプロセッサなどの日本語変換部で文字列を単語に区 切るためによく用いられる手法である。

【0011】ここで、単語辞書10には、ある方法で集 められた単語が登録されている。その方法は、まず、で きるだけ多くのサンプル文章の中から被験者(一人又は 思想をほぼ同一にする複数の人)が不快と感じる文章を 選別する。そして、その選別された文章(テキストデー タの場合はそのままで、また、印刷物の場合は文字認識 装置にかけて)を、例えば、上記の形態素解析などの手 10 法を用いて単語に分解し、その単語のうち出現頻度の多 いものを登録するというものである。したがって、上記 のフローチャートによれば、受信した電子メールの添付 ファイルに、被験者が不快と思う単語が多く含まれてい るほど、基準値を超えやすくなるから、当該添付ファイ ルの表示を拒否又は警告して、受取人の目に触れないよ うにすることができ、受取人の感情を害しないという効 果が得られる。

【0012】なお、本実施の形態では、単語辞書10の 感覚が異なる(登録者の不快が必ずしも受取人の不快に ならない) ことが有り得るが、この場合は、受取人自ら の手によって単語辞書 10の内容を編集できるようにし ておけばよい。

【0013】図3は同じくCPU2で実行する電子メー ルブログラムのフローチャートであり、特に、電子メー ルの送信部分を示す概略的なフローチャートである。と の図において、例えば、差出人による送信コマンドの実 行に応答して、電子メールの送信イベントが発生(S1 0) すると、まず、その電子メールの添付ファイルの有 30 無を判定(S11)する。そして、添付ファイルがない 場合は電子メールを所要の形式に加工した後、通信イン ターフェース7に送って送信(S12)する。一方、添 付ファイルがある場合は、その添付ファイルからテキス トデータを取り出し(S13)、受信時と同様の手法を 用いてテキストデータを単語に分解した後、単語ごとに ROM6の単語辞書10とマッチング(S14)をと り、一致した数をカウントする。そして、このカウント 値と基準値とを比較(S15)し、基準値を超えていな ければファイルの添付を許容(S16)して送信し、超 40 えている場合はファイルの添付を拒否(S17)する旨 を表示し、送信を中止するか添付ファイルなしで送信す るかを差出人の判断に委ねる。したがって、図3のフロ ーチャートによれば、送信対象の電子メールに添付する ファイルに、被験者(単語辞書10の登録者)が不快と 思う単語が多く含まれているほど、基準値を超えやすく なるから、当該添付ファイルの送信を拒否して、その電 子メールが他人(受取人)の目に触れないようにするこ とができ、他者の感情を害しないという効果が得られ る。

【0014】なお、以上の電子メール受信又は送信フロ ーチャート(図2又は図3)では、添付ファイルに含ま れる単語と単語辞書10とのマッチングをとっている が、これに限らない。電子メールの表題(Subject)や 本文に含まれる単語のマッチングをとってもよい。ま た、上述の電子メール受信フローチャート(図2)で は、受信電子メールごとにいちいち単語のマッチング処 理を行っているが、例えば、特定の差出人から不快感を 伴う電子メールが大量に送り付けられた場合は、メール の数だけマッチング処理が繰り返されるので、待ち時間 の増大を否めない。

【0015】図4及び図5に示すフローチャートは、と のような不都合を解決できるものである。すなわち、電 子メールを受信(S40)すると、上述のフローチャー ト(図2)と同様に、単語抽出(S41)とマッチング (S42) とを行った後、そのマッチングの結果得られ るカウント値をその電子メールの差出人のアドレスに関 連付けて差出人辞書11に格納する。そして、次通の電 子メールを受信(S50)すると、まず、その電子メー 登録者と受取人が別人であり、場合によっては、両者の 20 ルの差出人のアドレスと差出人辞書11とのマッチング (S51)をとり、そのマッチングの結果得られるカウ ント値と基準値とを比較(S52)し、基準値以下であ ればその電子メールの表示を許容し、基準値以上であれ ぱその電子メールの表示を拒否(S54)又は警告する というものである。したがって、このフローチャート (図4及び図5)によれば、一度、不快な電子メールを 送り付けた差出人のアドレスは、基準値を超える高いカ ウント値とともに差出人辞書11に登録されるから、以 降、単語のマッチングを行わず(したがって、待ち時間 の増大を伴わず) に、効率よく同一人からの電子メール を拒否又は警告することができるという有利な効果が得 **らわる**

> 【0016】なお、本発明の実施の形態は以上の例に限 定されない。その意図する範囲において様々な変形例を 包含するものである。例えば、図2、図3又は図5の基 準値を異なる値の複数とし、各基準値ごとに拒否や警告 のレベルを切り替えるようにしてもよい。また、以上の 例では、不快な電子メールの受信や送信を拒否している が、判読不能にしてもよく、例えば、該当する単語を伏 せ字にしたり、他の言語コードに置き換えたりしてもよ い。また、上述の単語マッチング処理(図2又は図3) において、カウント値を分析対象文章の文字数で正規化 することは好ましい。文字数が多い場合、単語の切り出 し誤差や偶然が積み重なる結果、不快でないにもかかわ らず、幾つかの単語がマッチして、不本意にもカウント 値がアップするからである。また、単語辞書10に単語 を登録する際、その単語を抽出したサンプル文書におけ るその単語の出現頻度も合わせて登録しておくと、マッ チングの際の精度が向上するから好ましい。

50 【0017】さらに、上記実施の形態の主要な機能(読

お取り手段、抽出手段、第1判定手段、カウンタ、第2 料定手段、単語辞書など)は、CPU2を含むハードウェア資産と、オペレーティングシステムや電子メールプログラムなどのソフトウェア資産との有機的結合によって機能的に実現されるものであり、電子メールプログラムを除く要素は汎用品(例えば、パーソナルコンピュータ)を利用できるから、本発明にとって欠くことのできない必須の事項は、実質的に、以上説明した電子メールプログラムに集約されているということがいえる。したがって、本発明は、以上説明した電子メールプログラムに集約されているということがいえる。したがって、本発明は、以上説明した電子メールプログラム 10のすべて又は要部を格納した、フロッピーディスク、MO、CD、ハードディスク、半導体メモリなどの記録媒体(それ自体が流通経路に乗るものはもちろん、ネットワーク上にあって記録内容だけを提供するものも含む)を包含するものである。

[0018]

【発明の効果】請求項1記載の発明によれば、主観上特定の思想に属する単語又は該思想を想起し得る単語(以下、特定単語)を含む文章を判定でき、かかる文章の閲覧や他者への配布を未然に防止できる。請求項2記載の 20発明によれば、特定単語を"多く"含む文章を判定でき、良好な判定精度でかかる文章の閲覧や他者への配布を未然に防止できる。請求項3記載の発明によれば、カ*

*ウンタの値を文章の文字数で正規化したので、文字数が 多い文章の判定精度を向上できる。請求項4記載の発明 によれば、ネットワークを介して送受信される電子メー ルのうち、特定単語を含むをものを排除でき、電子メー ルの健全的発展を図ることができる。請求項5記載の発 明によれば、電子メールの内容に応じて拒否又は警告の レベルを多段階に切り替えることができ、きめ細かな判 定を行うことができる。請求項6記載の発明によれば、 同一の差出人からの不快な電子メール受信を、短時間に 効率よく排除できる。

【図面の簡単な説明】

【図1】実施の形態の構成図である。

【図2】電子メールの受信フローチャートである。

【図3】電子メールの送信フローチャートである。

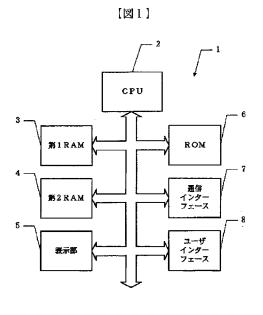
【図4】電子メールの差出人登録フローチャートであ ろ.

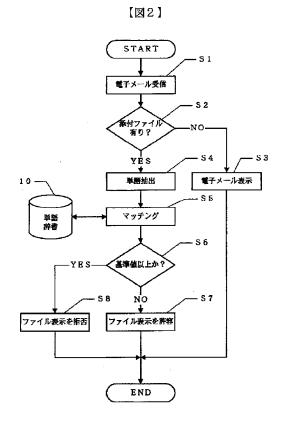
【図5】差出人による電子メール拒否フローチャートである。

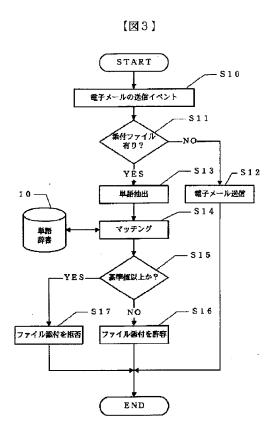
【符号の説明】

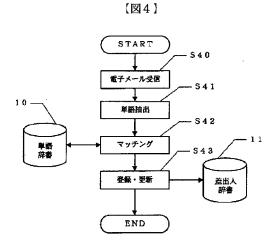
2 CPU (読み取り手段、抽出手段、第1判定手段、 カウンタ、第2判定手段)

10 単語辞書

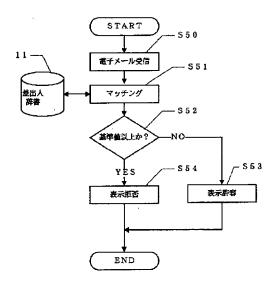








【図5】



フロントページの続き

(S1)Int.Cl.⁶ 識別記号 F I

H O 4 L 12/58 H O 4 L 11/20 1 O 1 B